



**Statistical Analysis and Adaptive Design Plan for the *PROSpect* Trial**
March 17, 2019

---

## 1.0 Introduction

This document describes the statistical analysis and adaptive design plan for *PROSpect* (Prone and Oscillation Pediatric Clinical Trial), a two-by-two factorial, response-adaptive, randomized controlled clinical trial of the efficacy of four positioning and ventilation strategies in severe pediatric acute respiratory distress syndrome (PARDS). Each of the four strategies is a combination of a positioning strategy and a ventilation strategy, namely:

- Supine positioning, conventional mechanical ventilation (CMV)
- Prone positioning, conventional mechanical ventilation
- Supine positioning, high-frequency oscillatory ventilation (HFOV)
- Prone positioning, high-frequency oscillatory ventilation.

The specific aims of *PROSpect* are to compare the effects of prone positioning with supine positioning and to compare the effects of HFOV with CMV. The four arms will be compared using the *PROSpect* trial primary endpoint, ventilator-free days (VFD) through day 28. VFD are defined as the number of days within 28 days that a patient is alive and free of mechanical ventilation. Patients will be assigned zero VFD if they remained intubated or died prior to day 28 without remaining extubated for more than 24 hours. VFD is the inverse equivalent of 28-day hospital mortality-adjusted duration of mechanical ventilation. Our hypothesis is that children with severe PARDS treated with prone positioning or HFOV will demonstrate more VFD.

The secondary aim of *PROSpect* is to compare the impact of these interventions on nonpulmonary organ failure-free days. Our hypothesis is that children with severe PARDS treated with prone positioning or HFOV will demonstrate more nonpulmonary organ failure-free days. Finally, we will explore the interaction effects of prone positioning with HFOV on VFD and investigate the impact of these interventions on 90-day in-hospital mortality and, among survivors, the duration of mechanical ventilation, PICU and hospital length of stay, and trajectory of post-PICU functional status and health-related quality of life.

The *PROSpect* trial utilizes a Bayesian adaptive design that includes multiple key features:

1. Adaptive sample size ranging from 400 to 1,000 patients
2. Response-adaptive randomization to favor well-performing arms
3. The ability to stop either of the two positioning strategies and/or either of the two ventilation strategies for efficacy
4. Stopping early for futility when it is unlikely that any more efficacy hypotheses can be resolved conclusively.

After providing an overview of the adaptive design, this document reviews the primary analyses for the primary endpoint and describes the details of the response-adaptive randomization and monitoring for efficacy and futility. Next, operating characteristics for this study design are presented under various scenarios, including sample size justification for the primary study aim. Finally, additional aspects of the statistical analysis plan, including analysis for the secondary and exploratory aims, are presented.

## 2.0 Adaptive Design Overview

The trial will have its first randomization update analysis after 400 patients are randomized and have been followed for 28 days. Subsequent randomization update analyses will occur after each additional 100 patients. The maximum total sample size is 1,000 (enrollment stops immediately with the randomization of the 1,000$^{th}$ patient, if not before).

At each randomization update analysis, the following decisions are possible, according to pre-specified rules described in Section 4.4 below:

- The trial may permanently drop either the supine positioning strategy or the prone positioning strategy, in which case all subsequently enrolled patients will be assigned the remaining positioning strategy.
- The trial may permanently drop either the CMV strategy or the HFOV strategy, in which case all subsequently enrolled patients will be assigned the remaining ventilation strategy.
- If one positioning strategy and one ventilation strategy have each been dropped, the trial stops early for efficacy.
- The trial may stop early for futility if it is unlikely that it will be able to make any more arm-dropping decisions, even with the full enrollment of 1,000 patients.
- Response-adaptive randomization probabilities will be updated for all arms that remain in the trial.

The primary analyses for the trial are conducted after all patients enrolled at the time of the stopping decision have been followed for 28 days. In particular, the trial may stop for apparent superiority of some combination of strategies, but one or more primary analyses may fall short of significance. Both primary analyses are based on the van Elteren test, a stratified version of the Wilcoxon rank-sum test. The positioning primary analysis compares VFD for patients assigned supine positioning with VFD for patients assigned prone positioning, stratifying by ventilation strategy. The ventilation primary analysis compares VFD for patients assigned CMV with VFD for patients assigned HFOV, stratifying by positioning strategy. The primary analyses are conducted at the one-sided 0.018 level as described in Section 3.4 below.

Decisions to permanently drop arms and to stop the trial are based on Bayesian predictive probabilities of significant primary analyses (see Section 4.3 for details). Response-adaptive randomization probabilities are based on the Bayesian posterior probabilities that each arm has the highest median VFD (i.e., lowest median duration of mechanical ventilation) among the arms (see Section 4.2 for details).

## 3.0 Study Population, Primary Endpoint, and Statistical Tests

### *3.1 Study Entry Criteria*

The trial enrolls pediatric patients (at least 2 weeks of age, at least 42 weeks post gestational age, and less than 18 years of age) intubated and mechanically ventilated with severe pediatric acute respiratory distress syndrome (PARDS) for less than 48 hours. More details on inclusion and exclusion criteria are in the study protocol.

### *3.2 Treatment Arms*

The trial begins by randomizing across the four treatment arms, initially with 1:1:1:1 randomization. None of the four arms is treated as a control arm. At some point in the trial, the trial may permanently drop two of the four arms (i.e., the two arms with supine positioning, the two arms with prone positioning, the two arms with CMV, or the two arms with HFOV). The permanent arm-dropping mechanism applies only to pairs of arms, but in principle, response-adaptive randomization may temporarily stop assigning positive probability to one or more of the arms.

### *3.3 Primary Endpoint*

The primary endpoint for this trial is ventilator-free days (VFD) through day 28. VFD are defined as the number of days within 28 days that a patient is alive and free of mechanical ventilation. Patients will be assigned zero VFD if they remained intubated or died prior to day 28 without remaining extubated for more than 24 hours. VFD is the inverse equivalent of 28-day hospital mortality-adjusted duration of mechanical ventilation. VFD will be recorded more precisely, to the minute.

### *3.4 Primary Analyses*

The trial has two primary analyses, one for positioning strategy and one for ventilation strategy. The analyses are based on the van Elteren test, a stratified version of the Wilcoxon rank-sum test. The primary analysis for positioning strategy treats the two ventilation strategies as strata and compares Supine/CMV patients with Prone/CMV patients, as well as comparing Supine/HFOV patients with Prone/HFOV patients. The positioning primary analysis will be considered statistically significant in favor of supine positioning if the one-sided van Elteren p-value is less than 0.018 for the null hypothesis that VFD for supine patients are no larger than VFD for prone patients. Similarly, if the one-sided p-value for the van Elteren test of the null hypothesis that VFD for prone patients are no larger than VFD for supine patients is less than 0.018, then the primary analysis for positioning will be considered statistically significant in favor of prone positioning. An analogous primary analysis will be conducted comparing the two ventilation strategies.

The threshold of 0.018 was chosen so that each of the four possible statistically significant results has approximately a 2.5% chance of occurring if all four arms are equivalent, and so that the probability of at least one significant result is less than 10%.

## 4.0 Prospectively Planned Randomization Update Analyses

The trial will have up to six randomization update analyses. The first one occurs after 400 patients are randomized and have been followed for 28 days. If the trial does not stop first, additional randomization update analyses will be conducted after 500, 600, 700, 800, and 900 patients are randomized and have been followed for 28 days.

### *4.1 Statistical Model*

Each of the four strategies has a probability of death $\pi_j$. The four $\pi_j$'s are assigned independent Beta(0.5, 0.5) priors, which is the standard Jeffreys prior. Using Supine/CMV as the base strategy, we define $\theta_{Prone}$, $\theta_{HFOV}$, and $\theta_{Interaction}$ as the effects of the positioning and ventilation strategies on duration of mechanical ventilation among those who do not experience death within 28 days, comparing these three strategies to Supine/CMV. For these Supine/CMV patients, we model the duration of mechanical ventilation as Gamma$(\alpha, \beta)$ with any values larger than 28 truncated to 28. The distribution of duration of mechanical ventilation for these patients for the other three strategies is defined similarly; we assume they all use the same shape parameter $\alpha$ but different rate parameters as follows:

- Supine/CMV: rate parameter $\beta$
- Prone/CMV: rate parameter $\beta \times \theta_{Prone}$
- Supine/HFOV: rate parameter $\beta \times \theta_{HFOV}$
- Prone/HFOV: rate parameter $\beta \times \theta_{Prone} \times \theta_{HFOV} \times \theta_{Interaction}$.

Here, $\alpha$ has a prior density proportional to $\alpha^{-1.5}$ on [1,100] and $\beta$ has an exponential prior distribution with mean 1/15. We place gamma priors on $\theta_{Prone}$, $\theta_{HFOV}$, and $\theta_{Interaction}$, each with prior mean 1. The gamma shape parameters are 3 for $\theta_{Prone}$ and $\theta_{HFOV}$ and 10 for $\theta_{Interaction}$. If all three $\theta$ parameters were equal to the prior means of 1, all four arms would have the same distribution for patients who do not experience death within 28 days. The larger shape parameter for $\theta_{Interaction}$ encourages the model to be approximately additive, unless it is clearly contradicted in the data.

### *4.2 Response-Adaptive Randomization*

For the first 400 randomized patients, allocation will be 1:1:1:1 among the four treatment arms, stratifying by age (<1; 1-7; 8-17 years) and by direct/indirect lung injury (6 strata in total). Stratification by age and type of lung injury will allow us to balance potentially important subgroups among the four intervention groups. Randomization will occur in permuted blocks with random block sizes of 4 and 8. Starting with the 400-patient randomization update analysis, the design shifts the randomization probabilities away from 1:1:1:1 based on evidence of efficacy. After estimating the statistical model defined in Section 4.1, we obtain, for each arm, the posterior probability that its median duration of mechanical ventilation is the lowest of the four arms. As described in Section 3.3, patients who experience death within 28 days or who are not extubated by day 28 will be assigned zero VFD or 28 days duration of mechanical ventilation. We define $M_{X/Y}$ to be the probability that strategy $X/Y$ has the lowest median duration of mechanical ventilation of the four arms, $SE_{X/Y}$ to be the posterior standard error of the median duration of mechanical ventilation for strategy $X/Y$, and $N_{X/Y}$ to be the number of

patients in the trial assigned to strategy $X/Y$. We construct new allocation probabilities beginning with defining:

$$r_{X/Y} = \frac{\sqrt{M_{X/Y} SE_{X/Y}}}{N_{X/Y}}.$$

These $r_{X/Y}$ values are normalized to sum to 1. If, after normalization, any value is under 5%, those values will be truncated to zero and the remaining $r_{X/Y}$ values renormalized. The renormalized $r_{X/Y}$'s are then used as the new randomization probabilities that will be used until the next randomization update analysis.

*4.3 Predictive Probabilities*

Decisions made as a result of randomization update analyses are based on Bayesian predictive probabilities using the statistical model defined in Section 4.1 with different assumptions about the remaining patients to be enrolled. Decisions about early efficacy are based on the current sample size, predicting the data sets that could be obtained if enrollment were to stop immediately and all enrolled patients were followed up for final primary endpoint data. Decisions about early futility are based on the maximum sample size, predicting data sets with the full 1,000 patients.

*4.4 Monitoring for Efficacy and Futility*

*4.4.1 Stopping Early for Efficacy (Permanent Dropping of a Pair of Arms)*

The trial aims to discover whether either of the two positioning strategies (supine or prone) is superior to the other, and whether either of the two ventilation strategies (CMV or HFOV) is superior to the other. The trial has the ability to permanently drop a pair of the arms corresponding to either a positioning strategy or a ventilation strategy if the evidence is strong enough for doing so, and it has the potential to stop altogether if it appears appropriate to choose both a positioning strategy and a ventilation strategy.

Suppose first that all four arms are still available (i.e., the design has not yet chosen either a positioning strategy or a ventilation strategy). The design decides whether to permanently drop a pair of arms as follows. The decision is based on the predictive probabilities of significant primary analyses assuming the trial stops enrollment immediately and follows up all enrolled patients until they have final data.

1. Draw a sample from the posterior distribution of the parameters of the statistical model defined in Section 4.1.
2. For each enrolled patient missing primary endpoint data, use the sampled parameters to draw a random primary endpoint value. For example, if a patient has been assigned to prone positioning and CMV, the patient is assigned a final endpoint of death (zero VFD) with probability $\pi_{Prone/CMV}$, and if the patient does not experience death within 28 days, their duration of mechanical ventilation is based on a gamma random variable with shape parameter $\alpha$ and rate parameter $\beta \times \theta_{Prone}$, as described in Section 4.1.
3. Evaluate the resulting final data set to determine whether any of the four one-sided van Elteren tests is significant at the 0.018 level.

4. Repeat this process for 100,000 samples from the posterior distribution, tabulating the fraction of simulated data sets where the primary analysis for positioning strategy is statistically significant in favor of supine positioning, the fraction of simulated data sets where the primary analysis for positioning strategy is statistically significant in favor of prone positioning, the fraction of simulated data sets where the primary analysis for ventilation strategy is statistically significant in favor of CMV, and the fraction of simulated data sets where the primary analysis for ventilation strategy is statistically significant in favor of HFOV. These are the four predictive probabilities used in early efficacy decisions.

If all four of the efficacy predictive probabilities are less than 0.95, then no pair of arms will be dropped. If one of them exceeds 0.95, the design permanently drops the corresponding pair of arms (e.g., if the primary analysis for ventilation strategy is significant in favor of CMV, the design drops the two HFOV arms). The response-adaptive randomization probabilities for the two remaining arms are then renormalized to sum to 1.

If a positioning strategy predictive probability exceeds 0.95 and a ventilation strategy predictive probability exceeds 0.95, the trial stops enrollment due to having chosen both a positioning strategy and a ventilation strategy.

If, at a previous randomization update analysis, the trial dropped two arms based on positioning strategy, the design computes only the two predictive probabilities for ventilation strategy. If one of the ventilation strategy predictive probabilities exceeds 0.95, the trial stops enrollment as it has now chosen both a positioning strategy and a ventilation strategy. Similarly, if the trial had previously dropped two arms based on ventilation strategy, then the design computes only the two predictive probabilities for positioning strategy.

Note that it is possible that the trial could stop enrollment due to efficacy calculations and yet ultimately fall short of statistical significance once final primary endpoint data are available for all enrolled patients. The predictive probability threshold of 0.95 is designed to help ensure this event is unlikely.

*4.4.2 Stopping Early for Futility*

The design also uses predictive probabilities to make futility decisions. In this trial without a specified control arm, futility means that it is unlikely that even with 1,000 patients the design will succeed in making any more decisions about either positioning strategy or ventilation strategy beyond the decisions already made.

First, assume that no efficacy decisions have been made and all four arms are still available. We calculate four predictive probabilities using steps similar to those in the efficacy calculations, with the additional step of predicting arm assignments for the patients yet to be enrolled.

1. Draw a sample from the posterior distribution of the parameters of the statistical model defined in Section 4.1.
2. Calculate the number of patients yet to be enrolled to achieve a final data set with 1,000 patients. Assign these patients treatment arms at random, according

to the current response-adaptive randomization probabilities calculated as described in Section 4.2, and assuming that these randomization probabilities do not change for the rest of the trial.
3. For patients currently enrolled and for the simulated patients from step 2, assign random primary endpoint values according to the sampled unknown parameters and the patients' real or simulated treatment assignments.
4. Evaluate the resulting final data set to see if any of the four van Elteren p-values are less than 0.018.
5. Repeat this process 100,000 times, obtaining four predictive probabilities.

If all four predictive probabilities are less than 10%, then the trial stops enrollment for futility.

If a previous randomization update analysis resulted in dropping two arms (e.g., the prone arms) for efficacy, then only two predictive probabilities are calculated (continuing the example, the two ventilation strategy predictive probabilities), and if both predictive probabilities are less than 10%, the trial stops enrollment for futility

If, at the current randomization update analysis, a pair of arms was dropped for efficacy, then another futility mechanism applies. Suppose the prone arms were dropped. If both remaining ventilation strategy predictive probabilities are less than 50%, and if they are not 10% larger than the corresponding (efficacy) predictive probabilities based on the currently enrolled patients, then the trial stops for efficacy of positioning strategy and futility of ventilation strategy. A similar process applies if a ventilation strategy pair of arms were dropped during the current randomization update analysis and positioning strategy predictive probabilities meet the same futility qualifications.

## 5.0 Operating Characteristics

In this section we present some operating characteristics for the design, estimated using simulation, for a limited number of scenarios. In the simulations for this document, we work with integer-valued duration of mechanical ventilation as an approximation, i.e., the statistical model assumes the time on the ventilator is a gamma random variable rounded down to the nearest integer. Operating characteristics for more precisely measured duration of mechanical ventilation should be similar, if anything slightly improved. First we describe assumptions common to all scenarios.

### *5.1 All Scenarios*

In all scenarios studied in this section, we make the same assumptions about the least effective arms(s). The following assumptions are based on patient data from the *RESTORE* (Randomized Evaluation of Sedation Titration for Respiratory Failure) trial, specifically VFD data from 712 patients who had severe PARDS with bilateral disease by the fourth day of intubation, were not intubated for asthma/reactive airways disease or bronchopulmonary dysplasia, and were not supported on ECMO. Ten percent of patients die, a further 4.2% of patients do not die but spend 28 days on the ventilator. The average VFD (including the patients who die) is 16 days (i.e., 12 days of mechanical ventilation). The distribution of VFD for patients who do not die is shown in Figure 1 below.

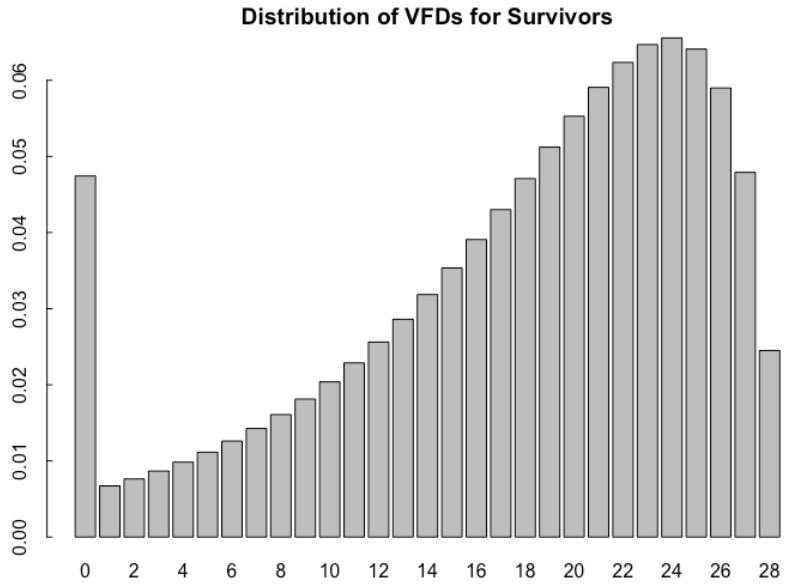


**Figure 1. Distribution of VFD for patients who do not die, for the least effective arm(s) in the scenarios simulated for operating characteristics**

The assumed accrual rate for scenarios simulated here is 1,000 patients over 3 years.

*5.2 Null Scenario*

For the null scenario, all four arms have the characteristics described in Section 5.1. We simulated 16,000 trials for this scenario. The first important set of operating characteristics is the probability of a significant primary analysis, shown in Table 1. Recall that primary analyses are conducted at the one-sided 0.018 level.

**Table 1. Type I Error Estimates**

| **Primary Analysis** | **Probability of Significance (Null Scenario)** |
|---|---|
| Supine > Prone | 0.0234 |
| Prone > Supine | 0.0258 |
| CMV > HFOV | 0.0243 |
| HFOV > CMV | 0.0268 |
| *Any of the four* | *0.0983* |

Each of the four possible primary analysis results has close to a 2.5% probability (the average is 2.51%). The probability of *any* significant primary analysis is slightly less than 10%.

Next we present summaries of sample size numbers for the simulated trials. The average sample size in the null scenario is 681 patients. The distribution of the time at which the trial ends is shown in Table 2. The most likely outcome is that the trial will stop at the 500-patient randomization update analysis; this happens about 20% of the time. Note that the trial will enroll slightly more than 500 patients in trials that stop at the 500-patient randomization update analysis.

**Table 2. Stopping Time Distributions in Null Scenario**

| **Stop Time** | 400 | 500 | 600 | 700 | 800 | 900 | 1,000 |
|---|---|---|---|---|---|---|---|
| **Probability** | 0.165 | 0.196 | 0.170 | 0.143 | 0.123 | 0.094 | 0.110 |

Table 3 contains average sample sizes for each arm and standard deviations of sample sizes. The average sample size for the trial as a whole is 681 patients in the null scenario, approximately 170 for each arm.

**Table 3. Sample Size Operating Characteristics for Null Scenario**

| **Arm** | **Mean (SD) Number of Patients** |
|---|---|
| Supine/CMV | 167 (61) |
| Prone/CMV | 174 (65) |
| Supine/HFOV | 170 (63) |
| Prone/HFOV | 170 (63) |
| *Total* | *681 (189)* |

Finally, we present summaries of how likely it is that the trial will permanently drop a pair of arms, and when. It is very rare to drop a pair of arms in the null scenario; there is only a 2% to 2.5% chance for each possible pair of arms. Table 4 shows these numbers; for example, there is a 0.7% chance that the design will drop the pair of prone arms at the 400-patient randomization update analysis.

**Table 4. Arm Dropping Probabilities in Null Scenario**

| **Dropped Arm Pair** | 400 | 500 | 600 | 700 | 800 | 900 | Never |
|---|---|---|---|---|---|---|---|
| Supine | 0.008 | 0.004 | 0.004 | 0.003 | 0.003 | 0.002 | 0.976 |
| Prone | 0.007 | 0.003 | 0.003 | 0.003 | 0.003 | 0.003 | 0.978 |
| CMV | 0.008 | 0.004 | 0.003 | 0.003 | 0.003 | 0.002 | 0.977 |
| HFOV | 0.007 | 0.004 | 0.004 | 0.003 | 0.003 | 0.002 | 0.977 |

### *5.3 Scenario 1: Positioning Strategy Effect Only (or, Separately, Ventilation Strategy Effect Only)*

In this section we consider scenarios where one positioning strategy is better than the other by 2 VFD on average, but the two ventilation strategies are equivalent. We also consider scenarios where one ventilation strategy is better than the other by 2 VFD, but the two positioning strategies are equivalent. The operating characteristics in this section are based on 1,000 simulations per scenario.

Table 5 shows the probabilities of possible primary analysis results in these scenarios. Here each column is a scenario, labeled by the better positioning strategy or better ventilation strategy, and each row is a potentially significant primary analysis. For example, when the prone position is better than the supine position but both ventilation strategies are equivalent, there is an 86.2% chance that the primary analysis for positioning strategy will be significant in favor of prone, zero chance it will be significant in favor of supine, and the probability of a significant result in favor of one ventilation strategy or the other is 0.016 + 0.026 = 0.042 (4.2%). The power numbers here range from 86% to 90%, while the average of the eight probabilities of a significant result in favor of a factor that is equally good as its opponent is 2.3%.

**Table 5. Primary Analysis Results for Scenarios with a Superior Positioning Strategy or a Superior Ventilation Strategy**

| **Superior Positioning** | **Supine** | **Prone** | **Both Equal** | **Both Equal** |
|---|---|---|---|---|
| **Superior Ventilation** | **Both Equal** | **Both Equal** | **CMV** | **HFOV** |
| Supine > Prone | 0.880 | 0.000 | 0.030 | 0.018 |
| Prone > Supine | 0.000 | 0.862 | 0.027 | 0.025 |
| CMV > HFOV | 0.021 | 0.016 | 0.883 | 0.000 |
| HFOV > CMV | 0.023 | 0.026 | 0.000 | 0.896 |

Rows are primary analyses and columns are scenarios.

Table 6 shows the distributions of stopping times for these scenarios. Each row is a scenario and each column is a possible time for the trial to stop.

**Table 6. Stopping Time Distributions in Scenarios with a Superior Positioning Strategy or a Superior Ventilation Strategy**

| **Stop Time** | 400 | 500 | 600 | 700 | 800 | 900 | 1,000 |
|---|---|---|---|---|---|---|---|
| Supine | 0.138 | 0.176 | 0.200 | 0.150 | 0.110 | 0.091 | 0.135 |
| Prone | 0.122 | 0.197 | 0.158 | 0.144 | 0.123 | 0.100 | 0.156 |
| CMV | 0.129 | 0.186 | 0.172 | 0.136 | 0.136 | 0.091 | 0.150 |
| HFOV | 0.155 | 0.205 | 0.149 | 0.132 | 0.128 | 0.097 | 0.134 |

Rows are scenarios and columns are possible times for the trial to stop.

Table 7 shows the means and standard deviations of sample sizes by arm for these scenarios. Each column is a scenario, each row is an arm in that scenario, and each cell contains the mean sample size and standard deviation. Average total sample sizes are about 700 for these scenarios. The two good arms have average sample sizes of around 223 and the two inferior arms have average sample sizes of about 128.

**Table 7. Sample Size Distributions in Scenarios with a Superior Positioning Strategy or a Superior Ventilation Strategy**

| **Arm** | **Supine** | **Prone** | **CMV** | **HFOV** |
|---|---|---|---|---|
| Supine/CMV | 216 (88) | 125 (31) | 218 (89) | 125 (32) |
| Prone/CMV | 130 (36) | 237 103) | 232 (98) | 130 (36) |
| Supine/HFOV | 223 (94) | 129 (36) | 129 (37) | 218 (93) |
| Prone/HFOV | 126 (33) | 218 (87) | 126 (34) | 219 (95) |
| *Total* | *695 (189)* | *709 (193)* | *705 (192)* | *692 (194)* |

Each column is a scenario and each row is an arm in that scenario. In each cell is the mean sample size (standard deviation).

Tables 8 through 11 show distributions of time to permanently drop pairs of arms. Each table is a scenario, each row is a pair of arms that can be dropped, and each column is a randomization update analysis at which the dropping decision can happen. Never in the 4,000 simulated trials did the design permanently drop a pair of arms corresponding to either superior factor.

**Table 8. Probabilities of Dropping Each Pair of Arms for the Scenario Where the Supine Arms are Superior**

| **Dropped Arm Pair** | 400 | 500 | 600 | 700 | 800 | 900 | Never |
|---|---|---|---|---|---|---|---|
| Supine | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| Prone | 0.408 | 0.147 | 0.121 | 0.065 | 0.055 | 0.049 | 0.155 |
| CMV | 0.007 | 0.004 | 0.002 | 0.002 | 0.000 | 0.001 | 0.984 |
| HFOV | 0.009 | 0.003 | 0.002 | 0.001 | 0.002 | 0.002 | 0.981 |

**Table 9. Probabilities of Dropping Each Pair of Arms for the Scenario Where the Prone Arms are Superior**

| **Dropped Arm Pair** | 400 | 500 | 600 | 700 | 800 | 900 | Never |
|---|---|---|---|---|---|---|---|
| Supine | 0.411 | 0.151 | 0.096 | 0.079 | 0.052 | 0.037 | 0.174 |
| Prone | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| CMV | 0.008 | 0.003 | 0.001 | 0.001 | 0.002 | 0.001 | 0.984 |
| HFOV | 0.003 | 0.003 | 0.002 | 0.004 | 0.001 | 0.001 | 0.986 |

**Table 10. Probabilities of Dropping Each Pair of Arms for the Scenario Where the CMV Arms are Superior**

| **Dropped Arm Pair** | 400 | 500 | 600 | 700 | 800 | 900 | Never |
|---|---|---|---|---|---|---|---|
| Supine | 0.012 | 0.003 | 0.003 | 0.004 | 0.003 | 0.000 | 0.975 |
| Prone | 0.013 | 0.003 | 0.004 | 0.001 | 0.001 | 0.001 | 0.977 |
| CMV | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| HFOV | 0.418 | 0.153 | 0.091 | 0.068 | 0.061 | 0.051 | 0.158 |

**Table 11. Probabilities of Dropping Each Pair of Arms for the Scenario Where the HFOV Arms are Superior**

| **Dropped Arm Pair** | 400 | 500 | 600 | 700 | 800 | 900 | Never |
|---|---|---|---|---|---|---|---|
| Supine | 0.007 | 0.002 | 0.000 | 0.001 | 0.005 | 0.005 | 0.980 |
| Prone | 0.008 | 0.003 | 0.001 | 0.000 | 0.001 | 0.003 | 0.984 |
| CMV | 0.444 | 0.143 | 0.078 | 0.081 | 0.079 | 0.046 | 0.129 |
| HFOV | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |

*5.4 Scenario 2: Positioning Strategy Effect and Ventilation Strategy Effect*

Next we present operating characteristics in scenarios where there is both a superior positioning strategy and a superior ventilation strategy. The superior positioning strategy leads to an average of 2 more VFD than the inferior positioning strategy, and the superior ventilation strategy leads to an average of 2 more VFD than the inferior ventilation strategy. The best arm, which is the combination of the superior positioning strategy and the superior ventilation strategy, is 4 days better than the worst arm, which is the combination of the inferior positioning strategy and the inferior ventilation strategy. The operating characteristics in this section are based on 1,000 simulations per scenario.

Table 12 shows the probability of significant primary analyses in these scenarios. For example, in the scenario where prone is the superior positioning strategy and HFOV is

the superior ventilation strategy, in 90.8% of simulated trials, the primary analysis for prone's superiority to supine was significant. Never in these 4,000 simulations did the trial conclude with a significant primary analysis in the direction of either the inferior positioning strategy or the inferior ventilation strategy. The probability of a significant primary analysis for the superior positioning strategy, which is about the same as the probability of a significant primary analysis for the superior ventilation strategy, is slightly above 90%. In other words, the design has slightly more than 90% power for positioning strategy, and slightly more than 90% power for ventilation strategy, in this scenario.

**Table 12. Primary Analysis Results for Scenarios with a Superior Positioning Strategy and a Superior Ventilation Strategy**

| **Superior Positioning** | **Supine** | **Prone** | **Supine** | **Prone** |
|---|---|---|---|---|
| **Superior Ventilation** | **CMV** | **CMV** | **HFOV** | **HFOV** |
| Supine > Prone | 0.909 | 0.000 | 0.907 | 0.000 |
| Prone > Supine | 0.000 | 0.907 | 0.000 | 0.908 |
| CMV > HFOV | 0.907 | 0.932 | 0.000 | 0.000 |
| HFOV > CMV | 0.000 | 0.000 | 0.924 | 0.906 |

Rows are primary analyses and columns are scenarios.

Table 13 shows the distributions of the time at which the trial stops for these four scenarios. The most likely time for the trial to stop is at the first randomization update analysis, and the second most likely possibility is getting all the way to 1,000 patients. The other possible stopping times are about equally likely.

**Table 13. Stopping Time Distributions for Scenarios with a Superior Positioning Strategy and a Superior Ventilation Strategy**

| **Stop Time** | 400 | 500 | 600 | 700 | 800 | 900 | 1,000 |
|---|---|---|---|---|---|---|---|
| Supine/CMV | 0.277 | 0.107 | 0.113 | 0.102 | 0.121 | 0.096 | 0.184 |
| Prone/CMV | 0.253 | 0.123 | 0.123 | 0.131 | 0.110 | 0.094 | 0.166 |
| Supine/HFOV | 0.264 | 0.121 | 0.091 | 0.129 | 0.090 | 0.118 | 0.187 |
| Prone/HFOV | 0.240 | 0.113 | 0.123 | 0.116 | 0.105 | 0.125 | 0.178 |

Rows are scenarios and columns are possible times for the trial to stop.

Table 14 summarizes sample size distributions by arm in these scenarios. For example, when the best arm is Prone/HFOV, an average of 108 patients is assigned to the Supine/CMV arm. The average sample sizes for the worst arms range from 108 to 110; recall that the first randomization update analysis occurs when 400 patients are randomized and have been followed for 28 days, so on average, more than 100 patients per arm will be in the trial before there is any chance to adapt. Average sample sizes for the best arm average 280 to 296, average sample sizes for arms with one superior factor and one inferior factor average about 148, and average total sample sizes are about 695.

**Table 14. Sample Size Distributions in Scenarios with a Superior Positioning Strategy and a Superior Ventilation Strategy**

| Arm | Supine/CMV | Prone/CMV | Supine/HFOV | Prone/HFOV |
|---|---|---|---|---|
| Supine/CMV | 280 (146) | 141 (53) | 150 (62) | 108 (12) |
| Prone/CMV | 152 (66) | 296 (155) | 110 (14) | 156 (69) |
| Supine/HFOV | 151 (62) | 109 (12) | 291 (150) | 147 (59) |
| Prone/HFOV | 109 (12) | 143 (54) | 145 (59) | 292 (145) |
| *Total* | *691 (220)* | *688 (212)* | *697 (220)* | *703 (215)* |

Each column is a scenario and each row is an arm in that scenario. In each cell is the mean sample size (standard deviation).

Tables 15 through 18 show distributions of time to permanently drop pairs of arms. There is almost a 50% chance that the inferior positioning strategy will be dropped at the first randomization update, and the same is true for the inferior ventilation strategy. Never in the 4,000 simulated trials did the design permanently drop a pair of arms corresponding to either the superior positioning strategy or the superior ventilation strategy.

**Table 15. Probabilities of Dropping Each Pair of Arms for the Scenario where Supine/CMV is the Best Arm**

| Dropped Arm Pair | 400 | 500 | 600 | 700 | 800 | 900 | Never |
|---|---|---|---|---|---|---|---|
| Supine | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| Prone | 0.484 | 0.105 | 0.089 | 0.064 | 0.067 | 0.045 | 0.146 |
| CMV | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| HFOV | 0.476 | 0.115 | 0.067 | 0.057 | 0.062 | 0.053 | 0.170 |

**Table 16. Probabilities of Dropping Each Pair of Arms for the Scenario Where Prone/CMV is the Best Arm**

| Dropped Arm Pair | 400 | 500 | 600 | 700 | 800 | 900 | Never |
|---|---|---|---|---|---|---|---|
| Supine | 0.467 | 0.122 | 0.105 | 0.065 | 0.057 | 0.046 | 0.138 |
| Prone | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| CMV | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| HFOV | 0.470 | 0.136 | 0.074 | 0.099 | 0.055 | 0.045 | 0.121 |

**Table 17. Probabilities of Dropping Each Pair of Arms for the Scenario Where Supine/HFOV is the Best Arm**

| Dropped Arm Pair | 400 | 500 | 600 | 700 | 800 | 900 | Never |
|---|---|---|---|---|---|---|---|
| Supine | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| Prone | 0.481 | 0.124 | 0.069 | 0.077 | 0.049 | 0.048 | 0.152 |
| CMV | 0.467 | 0.107 | 0.090 | 0.083 | 0.053 | 0.064 | 0.136 |
| HFOV | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |

**Table 18. Probabilities of Dropping Each Pair of Arms for the Scenario Where Prone/HFOV is the Best Arm**

| **Dropped Arm Pair** | 400 | 500 | 600 | 700 | 800 | 900 | Never |
|---|---|---|---|---|---|---|---|
| Supine | 0.463 | 0.124 | 0.085 | 0.069 | 0.053 | 0.060 | 0.145 |
| Prone | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| CMV | 0.459 | 0.111 | 0.081 | 0.054 | 0.057 | 0.060 | 0.146 |
| HFOV | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |

*5.5 Scenario 3: Non-Additive Model*

Finally we consider a more difficult set of scenarios, where there is a single superior arm that is 2 VFD better than the three other arms on average. In other words, the effects of positioning strategy and ventilation strategy are non-additive, and the first randomization update analysis happens when only about 100 patients assigned to an arm better than the worst arm have data. The primary analyses are also not ideally suited to this scenario: for example, if the supine/HFOV arm is the best arm, then supine is better than prone for HFOV patients, but supine is equivalent to prone for CMV patients. The operating characteristics in this section are based on 1,000 simulations per scenario.

Table 19 contains the distributions of primary analysis results. For each of the characteristics of the superior arm, the probability that it will be in a successful primary analysis is about 45%. It is just possible (it happened in 2 out of 4,000 simulated trials) that one of the characteristics of the superior arm will wind up on the wrong end of a successful primary analysis.

**Table 19. Primary Analysis Results for Scenarios with a Single Superior Arm**

| **Superior Arm** | Supine/CMV | Prone/CMV | Supine/HFOV | Prone/HFOV |
|---|---|---|---|---|
| Supine > Prone | 0.498 | 0.000 | 0.450 | 0.000 |
| Prone > Supine | 0.000 | 0.435 | 0.000 | 0.447 |
| CMV > HFOV | 0.450 | 0.442 | 0.001 | 0.000 |
| HFOV > CMV | 0.001 | 0.000 | 0.434 | 0.481 |

Table 20 shows the distributions of stopping times for these scenarios. Compared to the previous scenarios, these scenarios tend to stop later.

**Table 20. Stopping Time Distributions When There is a Single Superior Arm**

| **Stop Time** | 400 | 500 | 600 | 700 | 800 | 900 | 1,000 |
|---|---|---|---|---|---|---|---|
| Supine/CMV | 0.087 | 0.082 | 0.068 | 0.102 | 0.108 | 0.138 | 0.415 |
| Prone/CMV | 0.091 | 0.095 | 0.083 | 0.078 | 0.109 | 0.130 | 0.405 |
| Supine/HFOV | 0.080 | 0.084 | 0.090 | 0.096 | 0.108 | 0.149 | 0.393 |
| Prone/HFOV | 0.075 | 0.083 | 0.097 | 0.113 | 0.095 | 0.164 | 0.373 |

Table 21 shows the means and standard deviations of sample sizes by arm for these scenarios. Average total sample sizes are about 824 for these scenarios, larger than for previously studied scenarios. In each scenario, the good arm has a sample size of about 350, so on average, 43% of patients get the best arm. Even though all three of the inferior arms are equally ineffective, average sample sizes are smaller for the arm with nothing in common with the best arm, indicating that the additive model is having some effect on allocation.

**Table 21. Sample Size Distributions in Scenarios with a Single Superior Arm**

| Arm | Supine/CMV | Prone/CMV | Supine/HFOV | Prone/HFOV |
|---|---|---|---|---|
| Supine/CMV | 344 (136) | 158 (53) | 160 (51) | 142 (38) |
| Prone/CMV | 173 (58) | 356 (149) | 151 (45) | 170 (58) |
| Supine/HFOV | 168 (53) | 146 (42) | 351 (141) | 160 (52) |
| Prone/HFOV | 145 (41) | 161 (53) | 163 (53) | 349 (140) |
| *Total* | *828 (198)* | *821 (203)* | *824 (196)* | *821(194)* |

Each column is a scenario and each row is an arm in that scenario. In each cell is the mean sample size (standard deviation).

Tables 22 through 25 show distributions of time to permanently drop pairs of arms. Occasionally, on the order of once in every 1,000 simulated trials, a pair of arms including the best arm is dropped.

**Table 22. Probabilities of Dropping Each Pair of Arms for the Scenario Where the Supine/CMV Arm is Superior**

| Dropped Arm Pair | 400 | 500 | 600 | 700 | 800 | 900 | Never |
|---|---|---|---|---|---|---|---|
| Supine | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| Prone | 0.101 | 0.065 | 0.053 | 0.060 | 0.059 | 0.060 | 0.602 |
| CMV | 0 | 0 | 0 | 0.001 | 0 | 0 | 0.999 |
| HFOV | 0.066 | 0.055 | 0.069 | 0.059 | 0.049 | 0.055 | 0.647 |

**Table 23. Probabilities of Dropping Each Pair of Arms for the Scenario Where the Prone/CMV Arm is Superior**

| Dropped Arm Pair | 400 | 500 | 600 | 700 | 800 | 900 | Never |
|---|---|---|---|---|---|---|---|
| Supine | 0.092 | 0.072 | 0.058 | 0.048 | 0.058 | 0.040 | 0.057 |
| Prone | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| CMV | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| HFOV | 0.099 | 0.066 | 0.048 | 0.046 | 0.040 | 0.057 | 0.644 |

**Table 24. Probabilities of Dropping Each Pair of Arms for the Scenario Where the Supine/HFOV Arm is Superior**

| Dropped Arm Pair | 400 | 500 | 600 | 700 | 800 | 900 | Never |
|---|---|---|---|---|---|---|---|
| Supine | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| Prone | 0.088 | 0.053 | 0.063 | 0.049 | 0.054 | 0.056 | 0.637 |
| CMV | 0.085 | 0.066 | 0.069 | 0.052 | 0.037 | 0.050 | 0.641 |
| HFOV | 0.001 | 0.001 | 0 | 0 | 0 | 0 | 0.998 |

**Table 25. Probabilities of Dropping Each Pair of Arms for the Scenario Where the Prone/HFOV Arm is Superior**

| Dropped Arm Pair | 400 | 500 | 600 | 700 | 800 | 900 | Never |
|---|---|---|---|---|---|---|---|
| Supine | 0.102 | 0.064 | 0.060 | 0.057 | 0.058 | 0.051 | 0.608 |
| Prone | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| CMV | 0.084 | 0.064 | 0.065 | 0.062 | 0.045 | 0.052 | 0.628 |
| HFOV | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |

*5.6 Additional Considerations*

As an approximation, the simulations above worked with integer-valued VFD (and durations of mechanical ventilation), while VFD will be measured more precisely, to the minute, in the *PROSpect* trial. We expect that operating characteristics would be very similar for more precisely measured VFD. Finally, we will ensure balance in randomizations among the six age and lung injury strata using the method of Saville and Berry (Balanced covariates with response adaptive randomization, *Pharmaceut Statist.*, 2017). This method uses odds ratios to modify the new randomization probabilities described in Section 4.2 to obtain new stratum-specific randomization probabilities to balance the distribution of strata across treatment arms. Again, we expect that the operating characteristics would be very similar even with this balancing method.

## 6.0 Other Statistical Considerations

*6.1 Analysis Data Sets*

Data sets for DSMB reports, randomization update analyses, and final data analyses consist only of data for which all queries have been resolved.

**Intention-to-Treat Analysis Data Set:** The intention-to-treat data set consists of all randomized patients. Patients will be classified according to the treatment randomized regardless of actual treatment received. The ITT data set will be used for analysis of the primary outcome, including DSMB reports, randomization update analyses, and final data analyses. Missing data during the hospital stay is expected to be minimal, as patients have severe respiratory disease, and we expect minimal parental withdrawal during patient hospital stays. If the primary outcome is not known at the end of the study, the worst possible outcome (i.e., zero ventilator-free days) will be assigned.

**Per-Protocol Analysis Data Set:** The per-protocol data set consists of all randomized patients, except patients who never received the intervention, patients withdrawn from the protocol during the first 24 hours post-randomization by a clinician or parent/legal guardian, and patients whose parent/legal guardian withdrew full consent for the protocol and data collection. The per-protocol dataset will be used for analysis of all primary, secondary, and exploratory outcomes, including DSMB reports and final data analyses. Only patients included in the per-protocol data set will be eligible for follow-up.

*6.2 Further Analysis of the Primary Outcome*

In addition to the primary analyses described in Section 3.4, we will also evaluate possible interaction effects between the positioning and ventilation strategies, which will allow us to probe for potential differential effects when these two strategies are used concurrently. Although we may have low power to detect possible effect modification between positioning and ventilation strategies, we will explore for them using the statistical model in Section 4.1. If a significant interaction is found, a separate analysis will be conducted comparing all four combination strategies separately.

In addition to intention-to-treat analyses, analyses of the primary outcome will also be performed on a per-protocol basis. We will also explore adjustment for age group (<1; 1-7; 8-17 years) and lung injury type (direct; indirect) using proportional hazards

regression models, and we will make graphical comparisons using boxplots and Kaplan-Meier survival curves.

*6.3 Analysis of the Secondary Outcome*

Similar to the analysis of the primary outcome, analysis of the secondary outcome, nonpulmonary organ failure-free days, will also use stratified Wilcoxon rank-sum tests. Differences between positioning or ventilation strategies will be considered statistically significant if the two-sided p-value is <0.025. This analysis will be performed on a per-protocol basis. In addition, we will explore adjustment for age group and lung injury type using proportional hazards regression models.

*6.4 Analysis of the Exploratory Outcomes*

Analyses of exploratory outcomes will use logistic regression for binary outcomes (90-day in-hospital mortality), proportional hazards regression for time to event outcomes (durations of mechanical ventilation, PICU stay and hospital stay among survivors) and linear regression for continuous outcomes. For non-normal continuous outcomes, data transformations or nonparametric methods will be considered, as appropriate. These analyses will be performed on a per-protocol basis and will control for age group and lung injury type. Differences between positioning or ventilation strategies will be considered statistically significant if the two-sided p-value is <0.025. Careful assessment of the results from exploratory analyses will be made, though no formal multiple comparisons procedures are planned.

We will use appropriate methods for longitudinal outcomes, including random effects models or generalized estimating equations, to model repeated measures outcomes from the follow-up study, including functional status and health-related quality of life.

Throughout, descriptive statistics will be calculated, including means, standard deviations, medians, interquartile ranges, and ranges for continuous variables and frequency counts and percentages for categorical variables. Data will be examined for skewness, outliers, and systematic missing data. Residual analyses and model fit assessments will be performed to assess the appropriateness of modeling assumptions and check for outlying or overly influential observations.

*6.5 Additional Analyses*

PARDS is a complex disease having many causes and, among PARDS patients, responses to any intervention may be heterogeneous. The net benefit for an individual patient likely depends on the amount of potentially recruitable lung. Thus, we will perform a post-hoc analysis of responders, defined as patients who exhibit an increase in $PaO_2/FiO_2$ ratio of at least 20 or a decrease in oxygenation index [$OI = (FiO_2 \times \text{mean airway pressure} \times 100)/PaO_2$] of at least 10% within 24 hours of starting an intervention. In addition, we will tabulate the number of patients who switch to the reciprocal therapy (i.e., treatment failures).

We will also examine for time trends on outcome measures or treatment group effects (due to seasonal variation or learning effects) for primary and secondary outcome measures. If necessary, we will adjust for time in regression models. We do not expect effects of sex/gender or racial/ethnic group on outcome variables or treatment group

differences, but we will carefully examine for them. We will perform stratified analyses in subgroups and assess statistical interactions in the total sample. As necessary, we will present sex- and/or race-specific results. We will assess whether adjustment for site through the use of mixed effects or generalized estimating equations models or for region (North America, Europe, Australia, and Southeast Asia) through the use of fixed effects affects study inferences. We will also assess whether varying levels of protocol compliance result in varying levels of intervention effects using regression methods.

## 7.0 Conclusions

This document describes the goals and characteristics of the *PROSpect* two-by-two factorial response-adaptive trial design and presents estimates of its operating characteristics obtained via simulations. Overall, these simulation results confirm that the *PROSpect* trial, with up to 1,000 patients randomized in the two-by-two factorial, response-adaptive design, has high power to detect treatment group differences of 2 VFD. For example, when there is a superior positioning strategy but ventilation strategies are equivalent, we have approximately 88% power to detect the better strategy (Scenario 1, Table 5). When there is both a superior positioning strategy and a superior ventilation strategy, we have slightly greater than 90% power to detect each of the better strategies (Scenario 2, Table 12). Though we are not anticipating significant interaction effects between positioning and ventilation strategies, our study will allow an evaluation of potential synergistic or antagonistic effects. However, as expected, for a non-additive or interaction model the power is lower (Scenario 3, Table 19). However, in all cases, the response-adaptive design assigns more patients to strategies that are well-performing and allows early stopping for either efficacy or futility of positioning or ventilation strategy effects.

In addition, this document also provides details regarding the statistical analysis plans for the primary, secondary, and exploratory aims of the *PROSpect* trial, including intention-to-treat and per-protocol analyses and regression modeling. Overall, this clinical trial will provide the definitive evidence necessary for the field to consider a major change in clinical practice in the care of critically ill children with severe PARDS.